

Guía de Usuario: Encuesta Nacional sobre Niveles de Vida de los Hogares 2005

Luis Rubalcava

División de Economía

Centro de Investigación y Docencia Económicas A.C.

Graciela Teruel

Departamento de Economía

Universidad Iberoamericana

Octubre, 2007

Prefacio

Este documento describe los principales aspectos que permiten el análisis y manipulación de la información contenida en la ENNViH-2.

La Encuesta Nacional sobre Niveles de Vida de los Hogares (ENNViH) tiene como propósito crear una base de datos de tipo longitudinal y multitemática. Estas características permitirán por un lado, seguir a lo largo del tiempo a los individuos independientemente de sus decisiones de cambio de residencia, formación y/o desdoblamiento de hogares y por otro lado, contar en un solo instrumento con indicadores económicos, demográficos y de salud de la población mexicana. La primera ronda de la encuesta (ENNViH-1) se llevó a cabo en el año 2002 logrando encuestar a una muestra de 35 mil individuos en 150 localidades urbanas y rurales del país. La segunda ronda de la ENNViH (ENNViH-2) concluyó en el 2006 con tasas de recontacto cercanas al 90 por ciento a nivel hogar. En este segundo operativo se siguió y entrevistó a los miembros de hogares originalmente encuestados que entre 2002 y 2006 migraron a los Estados Unidos, alcanzando tasas de recontacto por arriba del 91 por ciento. En este contexto la ENNViH-2 provee evidencia estadística sobre el flujo de migración interna y hacia los Estados Unidos de la población y sobre la dinámica de bienestar de los mexicanos que se mueven dentro de nuestro país y hacia los Estados Unidos, independientemente si permanecen o regresan a territorio nacional y/o comunidad de origen. La ENNViH-2 esta constituida por una muestra cercana a 40 mil individuos debido al desdoblamiento asociado al ciclo de vida de los encuestados. El tercer levantamiento (ENNViH-3) está programado para el 2009.

La base de datos y la documentación de la ENNViH-1 y ENNViH-2 son de dominio público y se puede acceder a ellas en: www.ennvih-mxfls.org

La ENNViH-2 fue producto del esfuerzo conjunto de investigadores de la Universidad Iberoamericana (UIA), del Centro de Investigación y Docencia Económicas (CIDE), del Instituto Nacional de Salud Pública (INSP) y de la Universidad de California, Los Ángeles (UCLA).

La información, documentos y demás archivos de la ENNViH-2, serán de interés para la comunidad científica, para la discusión de la política social en México y la investigación comparativa con otros países.

Documentación para ENNViH-2:

1. Rubalcava, Luis y Teruel, Graciela (2006). "Encuesta Nacional sobre Niveles de Vida de los Hogares, Segunda Ronda", Documento de Trabajo, www.ennvih-mxfls.org. Se le solicita al usuario que utilice la base de datos de la ENNViH-2 que cite esta referencia.
 - 1.1. Diseño del proyecto
 - 1.2. Objetivo, trabajo de campo y tasas de respuesta para la encuesta, con énfasis en la ENNViH-2.
 - 1.3. Resultados de la ENNViH-1 y ENNViH-2

2. Berumen (2007). "Diseño de la Muestra". Descripción de la muestra de la ENNViH-2. Documento de Trabajo, Universidad Iberoamericana. Ciudad de México.
3. Cuestionarios de Hogares de la Segunda Ronda de la Encuesta Nacional sobre Niveles de Vida de los Hogares. Versiones en español e inglés.
4. Manuales de codificación de la encuesta de hogares de la segunda ronda de la Encuesta Nacional sobre Niveles de Vida de los Hogares. Versiones en español e inglés
 - 4.1. Descripción de las variables de los cuestionarios de la ENNViH-2 y su localización en los archivos de datos.
5. Cuestionarios comunitarios de la Segunda Ronda de la Encuesta Nacional sobre Niveles de Vida de los Hogares. Versiones en español e inglés
6. Manuales de codificación de la encuesta comunitaria de la segunda ronda de la Encuesta Nacional sobre Niveles de Vida de los Hogares. Versiones en español e inglés.
 - 6.1. Descripción de las variables de los cuestionarios de la ENNViH-2 y su localización en los archivos de datos.
7. Guía del usuario para la ENNViH-2.
 - 7.1. Descripción de la estructura de los archivos de la ENNViH y formato de los datos.
 - 7.2. Lineamientos para el uso de los datos, con énfasis en el cruce para el uso conjunto de las ENNViH-1 y ENNViH-2.

AGRADECIMIENTOS

Queremos expresar nuestro agradecimiento a todas las personas e instituciones que hicieron posible esta encuesta, en particular a:

- Las autoridades de la Universidad Iberoamericana (UIA), Ciudad de México y del Centro de Investigación y Docencia Económicas (CIDE)
- El Instituto Nacional de Salud Pública (INSP) por su colaboración en el diseño de los cuestionarios de salud y de dieta y la recolección de las medidas antropométricas de salud. Queremos agradecer muy especialmente el apoyo del Dr. Juan Rivera Dommarco, de la Mtra. Teresa Shama y del Dr. Salvador Villalpando.
- Berúmen y Asociados, por llevar a cabo el trabajo de campo y la captura de la información.
- El equipo de la ENNViH-2: Erika Arenas, Itzel Bello, Mariana Cabrera, Alejandra Castellón, Alejandro Celis, Karina Córdova, Guadalupe Correa, Celeste de la Huerta, Eder Figueroa, Antonio Flores, Carmina Flores, Cecil García, Delfino García, María José González, María Fernanda González, Marcelo González, Marisol González, Rogelio Guzmán, Ana Patricia Harari, Christian Herrera, Gabriela Langarica, Iliana Machuca, Michael Michaelson, Gerardo Molina, Alejandro Muratalla, Mónica Palacios, Juan Pablo Ramírez, Yaeli Reyes, Hugo Rojas, Vianey Rosales, Montserrat Rubio Luna, Claudia Ruiz, Luis Sánchez, Alicia Santana, María Fernanda Santiestebán, Amadeo Valdez, Liliana Vega.
- Las instituciones que contribuyeron al financiamiento de la ENNViH-2: Banamex, Fondos CONACYT-Sedesol, Instituto Mexicano del Seguro Social, National Institute of Child, Health and Development, Secretaría de Desarrollo Social, Secretaría de Salud, UC MEXUS y DHL.
- Finalmente agradecemos a todas y cada una de las personas que respondieron la ENNViH-2 por su interés, paciencia y participación. Sin su participación la ENNViH no hubiese sido posible.

Contenido	
Prefacio	ii
Agradecimientos	iv
1. Introducción	1
1.1 Colaboración Institucional	2
1.2 Planeación del Proyecto	3
1.3 Financiamiento	3
2. Diseño de la ENNViH-2	3
2.1 Diseño de los Cuestionarios de la ENNViH-2	3
2.2 Información del hogar y comunitaria	4
2.3 Confidencialidad	5
3. ENNViH-2: Elementos derivados de la ENNViH-1	6
3.1 Reentrevistas a los Hogares e Individuos de la ENNViH-1	6
3.2 Recontacto y seguimiento de los hogares e individuos en la ENNViH-2	6
3.3 Clasificación de Hogares Panel y Hogares Nuevos	7
3.4 Listas pre-impresas	7
4. Estructura de la ENNViH-2	8
4.1 Información Retrospectiva	11
4.2 Re-entrevistando a los servicios y comunidades de la ENNViH-2	12
5. Estructura de los archivos	13
5.1 Organización Básica de los archivos	13
5.2 Encuestas a Hogares	13
5.2.1 Identificadores y nivel de observación	14
5.3 Archivos de Cuestionarios de Hogares y de manuales de codificación	16
5.4 Archivos Comunitarios	16
6. Combinación de Archivos	17
6.1 Concatenando datos	17
6.2 Combinación de archivos a nivel individual, hogar y comunitario	18
6.2.1 Combinación de dos archivos a nivel individual	18
6.2.2 Cruzando dos archivos a nivel hogar	19
6.3 Combinación de archivos a nivel individual y hogar	19
6.4 Combinación de datos de hogares con datos comunitarios	20
7. Usando información de la ENNViH-2 con la ENNViH-1.	21
7.1 Diferencias en el identificado de la ENNViH-1 y la ENNViH-2	21

7.2 Pegando información de la ENNViH-1 y la ENNViH-2	22
7.2.1 Combinación de datos nivel hogar	22
7.2.2 Combinación de datos nivel individual	23
7.2.3 Combinación de datos nivel comunidad	23
8. Números de preguntas y nombre de las variables	24
8.1 Valores Missing y códigos especiales	24
8.2 Tipo de respuestas	24
9. Características especiales de la información de los hogares de la ENNViH-2	25
9.1 Información simétrica	25
9.2 Información duplicada	25
9.3 Relación entre los miembros del hogar	26
10. Limpieza de la información de la ENNViH-2	29
10.1 Sistema de búsqueda y proceso de limpieza “Look Ups”	29
10.2 Limpieza especial para respuestas abiertas	30
10.3 Sección de verificación	31
10.4 Verificación de los identificadores a través de los libros	32
10.5 Verificación de las unidades de medida	32
10.6 Creación de variables y archivos	32

1. Introducción

El objetivo principal de la ENNViH es generar una base de datos de propósitos múltiples y de corte longitudinal que abarque un periodo mínimo de 10 años. Esta base de datos permitirá entender mejor las transiciones económicas, demográficas y sociales por las que atraviese la población mexicana en el periodo de tiempo que dure el proyecto.

Investigadores de la UIA y del CIDE, en colaboración con el INEGI, llevaron a cabo en el 2002 el primer levantamiento de la encuesta (ENNViH-1). Este primer levantamiento, denominado la línea basal, recabó información detallada acerca del estado socioeconómico, demográfico, y de salud, de todos los miembros del hogar sobre una muestra de aproximadamente 8,400 viviendas localizadas a lo largo de todo el país. La encuesta es representativa a nivel nacional, urbano-rural y regional¹.

A nivel de individuos, la ENNViH-1 recaba información detallada de la escolaridad de cada miembro del hogar, historiales retrospectivos de migración, matrimonio, fecundidad y victimización sufrida fuera del hogar (asaltos, robos, secuestros); participación en la fuerza laboral; ingreso laboral de los adultos; transferencias monetarias y en especie; asignación de tiempo de los adultos y de los niños; créditos y préstamos; niveles de inversión en capital humano y toma de decisiones; percepciones del estado de salud, medidas objetivas del estado de salud (peso, talla, circunferencia de cintura y cadera, presión arterial y hemoglobina); salud reproductiva de todas las mujeres en el rango de edad fértil y uso de métodos anticonceptivos. A cada individuo—en rangos de edad apropiados--se le aplicó una batería de Matrices Progresivas de Raven, con el objeto de contar con una medida del estado cognoscitivo de los individuos.

Dado que el bienestar de las personas también depende del medio ambiente que los rodea (cantidad y calidad de infraestructura y servicios disponibles en la localidad), la ENNViH-1 realizó un levantamiento en campo paralelo, con entrevistas a centros de salud, escuelas, líderes comunitarios y establecimientos comerciales. Ello con el objetivo de capturar información cualitativa y cuantitativa, a nivel de localidad, sobre la infraestructura y servicios provistos en escuelas, centros de salud para conocer niveles actuales y pasados de infraestructura física y socioeconómica. Los instrumentos fueron aplicados en cada una de las localidades en que los encuestados vivían. Se recabó la siguiente información: historia y niveles actuales de la infraestructura económica y física de la comunidad (por ejemplo, salarios, disponibilidad de servicios públicos, infraestructura de comunicación y transporte, desastres naturales, crímenes, infraestructura de las escuelas y clínicas, oferta de servicios provistos en clínicas, precios de productos básicos, entre otros).

El diseño panel de la encuesta permite seguir a los mismos individuos y familias encuestadas en la línea basal a lo largo del tiempo, independientemente de su localización geográfica o movilidad económica. Continuar con la conformación del panel de la ENNViH enriquece considerablemente la información disponible para llevar a cabo análisis integral sobre el bienestar y sus cambios a través del tiempo en nuestro país.

1 Basada en la definición regional en el Plan Nacional de Desarrollo, México. 2000-2006.

Es por ello que investigadores del Centro de Investigación y Docencia Económicas (CIDE) y la Universidad Iberoamericana (UIA) unieron esfuerzos para llevar a cabo el segundo levantamiento de la ENNViH. La segunda ronda de la ENNViH (ENNViH-2) comenzó a mediados del 2005 y concluyó en el 2006. Como parte de la conformación del panel, el propósito fundamental de la ENNViH-2 fue re-entrevistar a todos los individuos y hogares que fueron encuestados en la línea basal (ENNViH-1) y a los individuos y hogares que debido al desdoblamiento se adicionaron a la muestra original. Dada la importancia del fenómeno migratorio que se vive en el país esta tarea significó uno de los retos más importantes para el proyecto. A pesar de ello, se pudieron obtener tasas de recontacto a nivel hogar cercanas al 90% de la muestra original.

Un reto adicional e inmenso representó el buscar, seguir y entrevistar a aquellos individuos que durante las rondas migraron a los Estados Unidos. El esfuerzo del equipo de la ENNViH aunado al apoyo recibido de la UCLA fueron piezas clave para lograr tasas de recontacto de migrantes a Estados Unidos por arriba del 91%. Hasta antes de la ENNViH-2, en México no existía una fuente de información que permitiera conocer, con detalle, las condiciones individuales y familiares de la población migrante antes y después de llevar a cabo su viaje. La ENNViH-2 provee información recopilada en ambos lados de la frontera lo que permite contar con datos adicionales sobre la dinámica de migración México-Estados Unidos. Las implicaciones de políticas públicas que puede significar el comprender mejor la compleja relación entre los emigrantes mexicanos y sus familiares en el hogar de origen serán trascendentales para el bienestar de la población.

La ENNViH-2 respeta el contenido, diseño y estructura de los cuestionarios de hogares y comunitarios de la ENNViH-1. Sin embargo, además de las secciones y preguntas que se incluyen en la línea basal, la ENNViH-2 incorpora nuevas secciones que contienen preguntas relacionadas con expectativas individuales, con preferencias intertemporales, altruismo y aceptación de riesgo. Asimismo, se incluyen preguntas nuevas sobre paternidad y guarderías.

1.1 Colaboración Institucional

Investigadores del Centro de Investigación y Docencia Económicas (CIDE) y la Universidad Iberoamericana (UIA) unieron esfuerzos para llevar a cabo el segundo levantamiento de la ENNViH (ENNViH-2). Investigadores de diversas instituciones participaron en el diseño de los nuevos módulos y preguntas que se integraron a la ENNViH-2. En particular se recibió ayuda especializada de expertos del departamento de Economía y Sociología de la Universidad de California, Los Ángeles (UCLA).

Para el trabajo de campo de la ENNViH-2 se firmaron convenios de colaboración con dos instituciones de prestigio. La empresa Berumen y Asociados, especialista en levantamiento de encuestas a hogares, estuvo a cargo de la parte operativa de la recolección de información a hogares. Por su parte, personal especializado del Instituto Nacional de Salud Pública estuvo a cargo de la recolección de las medidas antropométricas de salud y del diseño de los cuestionarios de dieta y de salud y de la recolección de la información a nivel comunitario.

1.2 Planeación del Proyecto

La planeación y dirección del proyecto es responsabilidad de Luis Rubalcava, Graciela Teruel, Duncan Thomas y Elizabeth Frankenberg.

1.3 Patrocinadores de la ENNViH-2

Las actividades previas al levantamiento de la información, recolección y publicación de la información fueron financiadas por diversas instituciones. Su contribución fue decisiva para completar el segundo levantamiento de la ENNViH. El equipo de la ENNViH-2 agradece a las siguientes instituciones:

- Banamex
- DHL
- Fondos CONACYT-SEDESOL
- Instituto Mexicano del Seguro Social (IMSS)
- National Institute of Child, Health and Development de los Estados Unidos (NICHD)
- Secretaria de Desarrollo Social (SEDESOL)
- Secretaría de Salud
- The University of California Institute for Mexico and the United States, (UC MEXUS)
- Universidad Iberoamericana, (UIA, FICSAC)

2. Diseño de la ENNViH-2

2.1 Diseño de los Cuestionarios de la ENNViH-2

La ENNViH-2 respeta el contenido, diseño y estructura de los cuestionarios de la ENNViH-1. Adicionalmente, la ENNViH-2 contiene módulos novedosos que recolectan información acerca de la formación de los hijos, expectativas individuales, preferencias inter-temporales, comportamiento sobre altruismo y aceptación de riesgo. Esto permitirá estudiar, por ejemplo, la relación entre expectativas y preferencias individuales, y el comportamiento de los integrantes del hogar en la toma de decisiones al interior de su hogar y su entorno comunitario. El diseño de estos nuevos módulos estuvo dirigido por especialistas de la UCLA, UIA y CIDE.

Además de los 3 libros de hogares: C (Libro de Control), I (Consumo del Hogar), II (Economía del Hogar) y de los 8 libros individuales : IIIA y IIIB (Características de los miembros del hogar de 15 años y más), IV (Salud reproductiva), V (Características de los miembros del hogar menores de 15 años), EA (Estado Cognoscitivo de los miembros del hogar de entre 13 y 65 años), EN (Estado Cognoscitivo de los miembros del Hogar de entre 5 y 12 años), el libro Proxy (adultos ausentes) y el libro S (salud) que integran a la ENNViH-1, la ENNViH-2 incorpora un nuevo libro; el libro D (libro de Dieta). El objetivo de este nuevo libro es conocer los patrones de dieta o de consumo de los miembros del hogar. Su diseño estuvo a cargo del Dr. Juan Rivera Dommarco del INSP.

Todos los módulos y preguntas adicionales a los cuestionarios de la ENNViH-1 fueron probados en distintas poblaciones de zonas urbanas, rurales e indígenas.

Una vez aceptado el funcionamiento de las nuevas preguntas se incorporaron a los cuestionarios de la ENNViH y se llevaron a cabo pilotajes de los libros finales. Al igual que la ENNViH-1, entrevistadores capacitados junto con personal especializado de Salud llevaron a cabo, en el estado de San Luis Potosí, una prueba piloto general que probara los protocolos del levantamiento de la información, la búsqueda de migrantes, la coordinación de los entrevistadores con el personal de salud, la edición de la información en campo y todas las actividades que implicaría el trabajo de campo.

2.2 Información del hogar y comunitaria.

La ENNViH-2 recolecta en una sola base de datos información detallada acerca de los hogares y las comunidades en las que habitan. La información de hogares incluye:

- Consumo y autoconsumo del hogar.
- Decisiones de ahorro, activos, deudas, ingreso laboral y no laboral, transferencias públicas y privadas y participación en los programas sociales de los hogares e individuos.
- Información retrospectiva de educación, interrupciones escolares y grados repetidos.
- Información retrospectiva de las decisiones laborales de cada miembro del hogar, incluyendo los niños.
- Asignación del tiempo.
- Información demográfica acerca de familiares no residentes del hogar.
- Información acerca de migración permanente y temporal.
- Información acerca del estado de salud de los individuos, enfermedades crónicas.
- Información acerca de la salud mental de los individuos.
- Información de medidas antropométricas para todos los miembros del hogar (peso, talla, circunferencia de cadera y cintura, presión arterial, nivel de hemoglobina, nivel de glucosa en la sangre, nivel de colesterol en la sangre y Proteína C Reactiva).
- Información retrospectiva detallada de la utilización de servicios de salud.
- Historia del embarazo, usos de métodos anticonceptivos, servicios médicos pos natales.
- Historia matrimonial y decisiones dentro del hogar.
- Incidencia de crímenes a nivel hogar e individual.
- Información detallada de negocios familiares y uso de tierra.
- Información acerca de situaciones de riesgo en la toma de decisiones.
- Información acerca del nivel de confianza en su entorno.
- Información sobre preferencias en el tiempo.
- Información acerca de la formación de los hijos.

- Información acerca de la dieta llevada a cabo por el hogar (tomando información representativa sobre la madre y un niño miembro del hogar).

Paralelamente al levantamiento de la información de hogares, la ENNViH-2 llevó a cabo la recolección de información a nivel comunitario. De forma similar a los cuestionarios de hogares, el diseño, estructura y contenido de los cuestionarios comunitarios es afín a los cuestionarios comunitarios de la ENNViH-1. Los cuestionarios de la ENNViH-2 comunitarios tienen el objetivo de completar la información a nivel hogar con información cuantitativa y cualitativa de la infraestructura de la localidad y precios con las características sociodemográficas de los hogares. En la ENNViH-2 sólo se recabó información sobre el libro de infraestructura y de precios.

La información a nivel comunidad incluye:

- Información general sobre la población que habita la localidad.
- Aspectos sociales de la localidad.
- Información acerca de desastres naturales.
- Historia de infraestructura.
- Asistencia social.
- Servicios educativos.
- Servicios médicos.
- Actividades comunitarias.
- Medios de transporte.
- Fuentes de agua.
- Electricidad.
- Oportunidades de crédito.
- Industria, empresas y fábricas.
- Club de migrantes.
- Precios de bienes y servicios en diferentes establecimientos mercantiles de la localidad.

2.3 Confidencialidad

Para asegurar la confidencialidad de los informantes de la ENNViH-2, toda la información que pudo haber sido usada para identificar a los informantes se eliminó de los archivos públicos. Esta información incluyó: nombres, direcciones, nombres de las localidades en las que vivieron en el pasado, nombres y direcciones de escuelas, centros de salud y proveedores de salud privados a los que asisten los miembros del hogar, entre otros.

Del mismo modo, con el fin de asegurar la confidencialidad de los servicios de la localidad el equipo de la ENNViH reemplazó los códigos de las localidades del INEGI con un código interno.

3. ENNViH-2: Elementos derivados de la ENNViH-1

En esta sección se presentan los elementos que fueron derivados de la ENNViH-1.

3.1 Reentrevistas a los Hogares e Individuos de la ENNViH-1.

El propósito de la ENNViH-2 es reentrevistar a todos los individuos y hogares que fueron encuestados en la línea basal (ENNViH-1) y a los individuos y hogares que debido al desdoblamiento se adicionaron a la muestra original. Para cada uno de los hogares originales (Hogar Panel) se creó una lista pre-impresa que contenía las características sociodemográficas más relevantes de los miembros que conformaban el hogar en el 2002: nombre, edad, fecha de nacimiento, sexo y relación con el jefe del hogar. La lista pre-impresa fue de gran utilidad durante la reentrevista porque indicaba al entrevistador acerca de las personas a las que tenía que entrevistar o sobre las que tenía que preguntar.

La importancia del fenómeno migratorio en México hizo que uno de los retos más importantes de la ENNViH-2 fuera localizar a los hogares e individuos que cambiaron su residencia entre el 2002 y 2005. De la muestra original de hogares e individuos encuestados en el 2002, el **6.88% y 11.73%** ya no vivían en el mismo domicilio. La tarea de seguimiento de hogares e individuos implicó desarrollar estrategias de búsqueda novedosas y complejas. En el caso de que el hogar panel ya no permaneciera en la dirección original los entrevistadores debían buscar e indagar acerca del paradero de ese hogar. Para ello, los entrevistadores contaban con información adicional de que les pudiera dar indicios de dónde encontrar al hogar. En el libro C de la ENNViH-1 se diseñó una sección (Sección recontacto) en la que se preguntaba por la dirección y teléfono de algún familiar o amigo muy cercano que en caso de que el hogar se mudara pudiera proporcionar información acerca de su paradero. Con esta información se creó un directorio de recontactos el cual se utilizó para identificar el paradero del hogar.

3.2. Recontacto y seguimiento de los hogares e individuos en la ENNViH-2

Para localizar a los hogares e individuos se llevaron a cabo una serie de operativos de recontacto con protocolos previamente establecidos, tales como verificación con los vecinos y personas del vecindario, preguntando si las personas del hogar realmente vivieron en el domicilio referido en el directorio. Durante la visita a la dirección original se trató de recabar información, acerca de la nueva localización del hogar con los vecinos e informantes clave de la comunidad.

El seguimiento de migrantes se llevó a cabo mediante tres estrategias: i) Recontactar a los hogares e individuos que se mudaron dentro de los 16 estados que conforman la muestra de la ENNViH, ii) recontactar a los hogares e individuos que se mudaron a estados que no pertenecen a la muestra de la ENNViH y iii) recontactar a los individuos u hogares que migraron a los Estados Unidos. Se formaron equipos especiales de

entrevistadores que hicieran posible estas labores. Con las estrategias establecidas se pudo localizar al 91% del total de individuos que migraron a los Estados Unidos y al 90% del total de hogares que conformaron la muestra original de la ENNViH.

3.3 Clasificación de Hogares Panel y Hogares nuevos.

Toda encuesta de tipo longitudinal se enfrenta a los problemas de atrición que puede generar, por un lado, la no respuesta y por otro, los cambios sociodemográficos que sufre la muestra original: muertes, nacimientos, casamientos, separaciones, etc. Las formas más comunes para hacer frente a la atrición de la encuesta panel son 1) actualizar la muestra de forma artificial ("refrescar" la muestra). Es decir, se adicionan hogares y/o individuos a la muestra original que no tienen que ver necesariamente con los hogares encuestados inicialmente y 2) actualizar la muestra de forma "natural" incorporando a la muestra original a todos aquellos individuos que ya sea por nacimientos o uniones ahora son considerados como integrantes del hogar.

En el caso de la ENNViH se decidió refrescar la muestra de forma natural. De esta manera, se define a un hogar o individuo panel como aquel que pertenece a la muestra original y que fue encuestado en el 2002. Un hogar panel puede modificar su estructura ya sea porque algún o algunos miembros panel salieron del hogar, porque otros murieron o porque llegaron nuevos miembros al hogar. En cualquier caso, la variable de interés son los individuos. Es decir, si miembros del hogar panel encuestados en el 2002 ya no pertenecen a este hogar en el 2005, se tendrá la tarea de buscarlos y encuestarlos. Si por el contrario, llegaron nuevos integrantes al hogar que no fueron encuestados en el 2002, pero actualmente se consideran miembros del hogar, serán clasificados como miembros nuevos y serán encuestados por primera vez.

Un hogar nuevo surge cuando uno o varios individuos encuestados en el 2002 salieron del hogar original y formaron un nuevo hogar. Por ejemplo, en el 2002 una persona que pertenecía a un hogar ENNViH salió de este hogar y se unió con otra persona (sea esta otra persona ENNViH o no ENNViH) esta unión conformará un nuevo hogar y se entrevistará al miembro panel y a todas personas que sean parte de este hogar. De esta manera, vamos a tener hogares nuevos e individuos nuevos que actualizarán la muestra original y resolverán los problemas de atrición que se pudieran presentar.

En todo hogar nuevo siempre deberá haber por lo menos un individuo panel. En un hogar panel puede haber o no individuos nuevos.

3.4 Listas pre-impresas

Con el objetivo de verificar la información obtenida en el primer levantamiento (ENNViH-1), corregir posibles errores y actualizar la información del hogar de forma más sencilla, se les proporcionó a los entrevistadores información pre-impresa de los miembros del hogar.

Las secciones de los cuestionarios para las que se crearon hojas pre-impresas son:

- **Lista Pre-impresa de miembros del Hogar (Libro C)**

Registra los nombres y características (sexo, edad, fecha de nacimiento, relación con el jefe del hogar, libros contestados en la primera ronda) de cada uno de los miembros que conformaban el hogar en el 2002.

- **Lista Pre-impresa de Centros, Unidades de Salud y escuelas a las que asisten los miembros del hogar (Libro C).**

Registra el nombre de los centros de salud, proveedores de salud privados y escuelas a los que asistían los miembros del hogar en el 2002 y que conformaron la muestra comunitaria de la ENNViH-1.

- **Lista Pre-impresa de la información de hijos, hermanos y otras personas no residentes del hogar (Libro IIIB).**

Registra la información de los hijos, hermanos y personas no residentes en el hogar, que los miembros del hogar proporcionaron en el 2002 correspondiente a la sección de transferencias del Libro IIIB.

Las listas pre-impresas sólo se utilizaron para el caso de hogares panel.

4. Estructura de la ENNViH-2.

La ENNViH-2 sigue la misma estructura de la ENNViH-1. En general, los mismos libros y las mismas preguntas fueron aplicados. Sin embargo, se adicionaron algunas preguntas y se quitaron otras. El siguiente cuadro muestra la estructura de los libros de la ENNViH-1 y ENNViH-2 y el informante adecuado para cada libro. La única diferencia en cuanto a los libros se refiere es que en la ENNViH-2 se adiciona el libro de Dieta, el cual no existía en el 2002. Este nuevo libro es respondido por la persona encargada de la preparación de los alimentos y por un niño miembro del hogar de entre 5 y 11 años. Respecto al informante adecuado, para los miembros panel va a ser el mismo que en el 2002.

Tabla 1. Estructura de los libros de hogares e informante adecuado ENNViH-1 y ENNViH-2			
Libro ENNViH-1	Informante ENNViH-1	Libro ENNViH-2	Informante ENNViH-2
Hogares			
C	MH* Mayor de 18 años (Preferentemente el jefe del hogar)	C	Mismo que en ENNViH-1
I	MH Mayor de 18 años (preferentemente el encargado del consumo/gasto del hogar)	I	Mismo que en ENNViH-1
II	MH Mayor de 18 años (Preferentemente el jefe del hogar)	II	Mismo que en ENNViH-1
Individuales			
IIIA	Todos los MH de 15 años y más	IIIA	Mismo que en ENNViH-1
IIIB	Todos los MH de 15 años y más	IIIB	Mismo que en ENNViH-1
IV	Todas las mujeres MH de entre 14 y 49 años de edad	IV	Mismo que en ENNViH-1

V	Todos los niños MH menores de 15 años	V	Mismo que en ENNViH-1
EA	Todos los MH de entre 13 y 65 años de edad	EA	Mismo que en ENNViH-1
EN	Todos los MH de entre 5 y 12 años de edad	EN	Mismo que en ENNViH-1
PROXY	Algún MH mayor de 18 años que conozca las características del miembro del hogar ausente (1 libro por cada miembro ausente)	PROXY	Mismo que en ENNViH-1
SALUD	Todos los MH	SALUD	Mismo que en ENNViH-1
-		DIETA	La persona MH encargada de preparar los alimentos y un niños MH de entre 5 y 11 años.

* MH: Miembro del Hogar

La siguiente tabla muestra las diferencias que existen entre la ENNViH-1 y ENNViH-2 en términos de preguntas y secciones de los cuestionarios. Las diferencias entre las primeras dos rondas de la encuesta se centran en las preguntas referentes a riesgo y confianza incluidas en el libro IIIB y libro II y a las referentes a formación de los hijos del libro V. Como parte de las técnicas de seguimiento y clasificación de individuos y hogares nuevos y paneles se adicionaron algunas preguntas en la sección LS del libro C.

Tabla 2. Diferencias entre preguntas y secciones ENNViH-1 y ENNViH-2			
Libro	ENNViH-1 Secciones	ENNViH-2 Secciones	Diferencias ENNViH-1 y ENNViH-2
C	DF (Definición)	DF (Definición)	Se adiciona la pregunta DF02 Para verificar si el hogar es panel o nuevo. Si es panel se deberá usar la hoja pre-impresa de la sección LS del libro C.
	LS (Lista de los miembros del hogar)	LS (Lista de los miembros del hogar)	Se adicionan las preguntas LS001-LS004 para asegurar que todos los miembros del hogar sean listados. Se adicionan las preguntas LS19-LS19f para verificar aspectos de miembros del hogar ausentes. Se adiciona la sección Lista de Direcciones
	CV (Características de la vivienda)	CV (Características de la vivienda)	Ninguna
	RC (Información de recontacto)	RC (Información de recontacto)	Ninguna
	SP (Salud y planificación familiar)	SP (Salud y planificación familiar)	Ninguna
	EH (Escolaridad de los MH)	EH (Escolaridad de los MH)	Ninguna
	CVO (Observación directa de las características de la vivienda)	CVO (Observación directa de las características de la vivienda)	Ninguna
	NE (notas de las secciones de la entrevista)	NE (notas de las secciones de la entrevista)	Ninguna
I	CS (consumo)	CS (consumo)	Ninguna
	NE (notas de las secciones de la entrevista)	NE (notas de las secciones de la entrevista)	Ninguna
II	SU (suelos)	SU (suelos)	Ninguna

	INR (Ingresos rurales)	INR (Ingresos rurales)	Ninguna
	NNA (Negocios no agrícolas)	NNA (Negocios no agrícolas)	Ninguna
	AH (Activos del Hogar)	AH (Activos del Hogar)	Ninguna
	CRH (Crédito del Hogar)	CRH (Crédito del Hogar)	Ninguna
	IN (Ingreso no laboral del hogar)	IN (Ingreso no laboral del hogar)	Ninguna
	SE (Shocks económicos del hogar)	SE (Shocks económicos del hogar)	Ninguna
	VLH (Violencia y victimización del Hogar)	VLH (Violencia y victimización del Hogar)	Se adicionaron los incisos de la k a la X de la pregunta VLH que captan cuestiones de confianza y altruismo.
	NE (Notas de las secciones de la entrevista)	NE (Notas de las secciones de la entrevista)	Ninguna
IIIA	ED (Educación)	ED (Educación)	Ninguna
	IE (interrupciones escolares)	IE (interrupciones escolares)	Ninguna
	AH (Activos del hogar)	AH (Activos del hogar)	Ninguna
	IIN (ingreso individual no laboral)	IIN (ingreso individual no laboral)	Ninguna
	HM (historia matrimonial)	HM (historia matrimonial)	Ninguna
	DH (decisiones del hogar)	DH (decisiones del hogar)	Ninguna
	MG (migración permanente)	MG (migración permanente)	Ninguna
	MT (migración temporal)	MT (migración temporal)	Ninguna
	TB (trabajo)	TB (trabajo)	Ninguna
	ATA (asignación del tiempo)	ATA (asignación del tiempo)	Ninguna
	SHI (Shocks Individuales)	SHI (Shocks Individuales)	Se adicionaron preguntas sobre expectativas.
	VLI (violencia y victimización del hogar)	VLI (violencia y victimización del hogar)	Ninguna
	NE (Notas de las secciones de la entrevista)	NE (Notas de las secciones de la entrevista)	Ninguna
IIIB	GH (Gustos y hábitos)		Ninguna
	-	EE (Expectativas)	Se creó la sección EE (Expectativas)
	-	RG (Riesgo)	Nueva sección en la ENNViH-2
	-	CO (Confianza)	Nueva sección en la ENNViH-2
	ES (estado de salud)	ES (estado de salud)	Ninguna
	SM (Estado de ánimo)	SM (Estado de ánimo)	Ninguna
	EC (Enfermedades crónicas)	EC (Enfermedades crónicas)	Ninguna
	ATS (Autotratamiento)	ATS (Autotratamiento)	Ninguna
	CE (Consulta externa)	CE (Consulta externa)	Ninguna
	HS (Utilización de servicios de hospitalización)	HS (Utilización de servicios de hospitalización)	Ninguna
	CA (condición de aseguramiento)	CA (condición de aseguramiento)	Ninguna
	RE (Recontacto de parientes en Estados Unidos)	RE (Recontacto de parientes en Estados Unidos)	Ninguna
	CR (crédito)	CR (crédito)	Ninguna
	PR (preferencias)	PR (preferencias)	Ninguna
	-	FH (Formación de los hijos)	Nueva sección en la ENNViH-2
	TP (Transferencias de padres no residentes)	TP (Transferencias de padres no residentes)	Ninguna
	TH (transferencias de hermanos no residentes)	TH (transferencias de hermanos no residentes)	Ninguna
	THI (transferencias de hijos no residentes)	THI (transferencias de hijos no residentes)	Ninguna
	TO (transferencias de otras personas no residentes)	TO (transferencias de otras personas no residentes)	Ninguna
	NE (notas de las sesiones de la entrevista)	NE (notas de las sesiones de la entrevista)	Ninguna

IV	RES (resumen del embarazo)	RES (resumen del embarazo)	Ninguna
	HE (historia del embarazo)	HE (historia del embarazo)	Ninguna
	AC (Anticoncepción)	AC (Anticoncepción)	Ninguna
	NE (notas de las sesiones de la entrevista)	NE (notas de las sesiones de la entrevista)	Ninguna
V	EDN (Educación del niño)	EDN (Educación del niño)	Ninguna
	EMN (empleo del niño)	EMN (empleo del niño)	Ninguna
	ATN (Asignación del tiempo del niño)	ATN (Asignación del tiempo del niño)	Ninguna
	CEN (Utilización de consulta externa del niño)	CEN (Utilización de consulta externa del niño)	Ninguna
	ESN (Estado de salud del niño)	ESN (Estado de salud del niño)	Ninguna
	HSN (Utilización de servicios de hospitalización del niño)	HSN (Utilización de servicios de hospitalización del niño)	Ninguna
	AUTN (Autotratamiento del niño)	AUTN (Autotratamiento del niño)	Ninguna
	VAC (Vacunación del niño)	VAC (Vacunación del niño)	Ninguna
	-	FH(formación de los hijos)	Nueva sección en la ENNViH-2
	NE (notas de las sesiones de la entrevista)	NE (notas de las sesiones de la entrevista)	Ninguna
S	SA (medidas de salud)		Se incorporaron las medidas de glucosa, colesterol y proteína C reactiva
	NE (notas de sesiones de la entrevista)		
D	-	CA (consumo de alimentos)	Nuevo libro en la ENNViH-2
	-	NE (Notas de las sesiones de la entrevista)	Nuevo libro en la ENNViH-2
EA	ECA	ECA	Ninguna
	NE (Notas de las sesiones de la entrevista)	NE (Notas de las sesiones de la entrevista)	Ninguna
EN	ECN	ECN	Ninguna
	NE (Notas de las sesiones de la entrevista)	NE (Notas de las sesiones de la entrevista)	Ninguna

4.1 Información Retrospectiva

Existen algunos módulos en los libros IIIA, IIIB y IV que fueron diseñados para recolectar información retrospectiva de los miembros del hogar. Ejemplo de estos módulos son, educación, historia matrimonial, migración y participación en la fuerza laboral. Los individuos que respondieron estos módulos en la ENNViH-1 (persona panel) no tuvieron que volver a responder el historial completo de estos módulos, únicamente actualizaron los historiales respondiendo a partir del año 2002 y hasta la fecha de la entrevista en la ENNViH-2. La clasificación del individuo como panel o nuevo de acuerdo a los libros que respondió en el 2002 la tiene el entrevistador en la hoja pre-impresa.

La ENNViH-2 generalmente recaba menos información en el caso de ser panel respecto a si se es nuevo. Los cuestionarios fueron estructurados 1) para recolectar la misma información retrospectiva de los individuos nuevos y panel y ii) para los panel sólo se actualiza la información recabada en la ENNViH-1, es decir, se pregunta acerca del tema en cuestión a partir del año 2002. De esta manera, tenemos información retrospectiva completa para los individuos panel de la ENNViH-2

La tabla 4 resume las diferencias en la información recolectada para los individuos panel y nuevos en los módulos retrospectivos y sus implicaciones al crear el historial completo del panel.

Tabla 3. Información retrospectiva ENNViH-1 y ENNViH-2

Sección	Informantes Nuevos	Informantes Panel	Historial Completo para Informantes Panel
IIIA-ED (Educación)	Último nivel de educación al que asiste/asistió	Último nivel de educación al que asiste/asistió a partir del ciclo 2001/2002.	Usar la información del módulo de educación de ENNViH-1 para la escolaridad antes del 2002. La escolaridad entre 2002 y 2005 es reportada en la ENNViH-1 y ENNViH-2
IIIA-ED (Grados repetidos)	Grados repetidos desde la primaria	Grados repetidos a partir del 2002	Usar la información del módulo de educación de ENNViH-1 para la escolaridad antes del 2002. La escolaridad entre 2002 y 2005 es reportada en la ENNViH-1 y ENNViH-2
IIIA-HM (Historia Matrimonial)	Historia matrimonial completa	Historia matrimonial a partir del 2002	Para todos los entrevistados que no se han casado/unido antes de 2001, ENNViH-2 da completa información de la historia matrimonial. Los datos sobre uniones/matrimonios que terminaron antes de 2001 están en la ENNViH-2
IIIA-MG (Migración Permanente)	Todas las preguntas desde que nació (Lugar donde nació y todos los movimientos después de los 12 años)	Todos los movimientos después de los 12 años.	Usar la ENNViH-1 para el lugar de nacimiento y residencia en los 12 años.
IIIA-TB (Trabajo) Verificar TB54x	Preguntas relacionadas con el último trabajo, el trabajo del 2003 y su primer trabajo	Preguntas relacionadas con el último trabajo y el trabajo del 2003	Para los entrevistados panel información adicional puede encontrarse en la ENNViH-1
IV-RES (Historia del Embarazo)	Historial de embarazos	Historial de embarazos después del 2001	Usar la ENNViH-1 para el módulo de historia del embarazo para los nacimientos ocurridos antes del 2001 Para los nacimientos ocurridos después del 2001 usar la ENNViH-2

4.2. Re-entrevistando a los servicios y comunidades de la ENNViH-2

Como se mencionó anteriormente, uno de los principales objetivos de la ENNViH-2 es conocer la evolución a través del tiempo de los factores que determinan el bienestar de los individuos. En este sentido, la ENNViH-2 pretende volver a entrevistar a todos los hogares, individuos, localidades y establecimientos mercantiles que fueron encuestados en la ENNViH-1.

La ENNViH-2 contiene información acerca de 300 localidades del país en donde habitan los individuos que conforman la encuesta.

5. Estructura de los archivos

Esta sección describe la organización, nombres y otras características de los archivos de datos para facilitar su análisis. Información adicional acerca de los archivos de datos se encuentra en los cuestionarios y manuales de codificación. Para un mejor análisis, cada página de los cuestionarios de hogares y comunitarios incluye los nombres de los archivos que contiene la información de esa página. Existe un manual de codificación para cada cuestionario, el cual describe el contenido de los archivos de los datos para ese cuestionario, así como el nivel de observación representada.

5.1. Organización Básica de los archivos.

Los archivos que contienen los datos de los cuestionarios de hogares y comunitarios se encuentran disponibles en el formato de Stata.

Los archivos de datos de la ENNViH-2 están organizados en tres grupos principales:

- Documentación descriptiva de la segunda ronda de la Encuesta Nacional Sobre Niveles de Vida de los Hogares (ENNViH-2)
- Archivos de datos de la encuesta de Hogares
- Archivos de datos de la encuesta Comunitaria

5.2. Encuestas a Hogares

Los archivos de la encuesta de hogares incluyen la base de datos, los cuestionarios y los manuales de codificación. La base de datos esta disponible en el formato de Stata 8.0. Para aquellos usuarios interesados en trabajar con otro tipo de formato se recomienda utilizar el software Stata Transfer. Este software permite convertir los archivos que están en el formato Stata a otros formatos como ASCII, dBase, Excel, SAS, SPSS, Statistica y SYSTAT.

Archivos de datos de las encuestas a Hogares:

- Cada archivo zipeado contiene los 12 libros de la encuesta a Hogares. Las letras hh son usadas para indicar que se refiere a hogar; los números 05 se refieren al año 2005 y dta indica que los archivos de datos se encuentran en el formato de Stata.
- Son doce los archivos que constituyen la encuesta de hogares. Estos archivos están expresados de la siguiente manera: hh05dta_byy, un archivo para cada módulo de cada libro de hogares; donde dta indica que este archivo corresponde a los datos y yy equivale al nombre del libro (c, 1, 2, 3A, 3B, 4, 5, S, D, EA, EN y X para el libro Proxy)
- Para todos los archivos de datos de la encuesta de hogares, existe su correspondiente archivo de cuestionario y de manual de codificación.

La nomenclatura de los archivos es muy fácil de entender. Los primeros caracteres están asociados con el cuestionario del libro al que pertenece, seguido por los caracteres que identifican el módulo específico y el número que se encuentra al final denota si la base de datos del modulo cuenta con múltiples archivos.

$$\frac{\text{xxx_xxx}}{\text{Libro_Módulo}} \quad \frac{\text{X}}{\text{Secuencia del archivo}}$$

Por ejemplo, el módulo de migración permanente (MG) del libro 3A, el nombre `iiia_mg` significa que el archivo contiene información del libro 3A, módulo MG y es la primera base de datos. El nombre `iiia_mg1`, denota el segundo archivo de la información del libro 3A, módulo MG.

5.2.1 identificadores y nivel de observación

Los datos de la encuesta se han organizado de la mejor manera posible para que el nivel de observación dentro de cada archivo sea a nivel del hogar o individual. Si el nivel de observación es a nivel hogar, la variable folio identifica una sola observación. Si el nivel es individual, el folio y ls son requeridas para poder identificar a una persona.

En la ENNViH-2, el folio es una variable conformada por 8 dígitos con el siguiente significado:

$$\frac{\text{x x x x x x}}{\text{Folio 2002}} \quad \frac{\text{x x}}{\text{original / nuevo}}$$

Los dos últimos dígitos, 00 es designado para los hogares originales de la ENNViH-1 que no se mueven. Para los hogares nuevos, los primeros 6 dígitos son iguales al folio original del cual se desprendió el individuo del nuevo hogar y los últimos dos dígitos se conforman utilizando el ls asignado en la ENNViH-1 al individuo que se mudó y que formó este nuevo hogar.

El ls en la ENNViH-2 para los individuos que se encuentran en hogares paneles es el mismo que el de la ENNViH-1. Si hay individuos nuevos que se adicionaron al hogar, el ls sigue su consecutivo, es decir, si en la ENNViH-1 había tres miembros y ocupaban el 01, 02 y 03, el nuevo miembro ocupará en la ENNViH-2 el 04, independientemente que alguno de los tres anteriores haya dejado el hogar o se haya muerto.

Los individuos que se encuentran en hogares nuevos y que son miembros originales no van a tener asignado el mismo ls que tenían en la ENNViH-1.

Por ejemplo, para el hogar con el folio "00001000" en la ENNViH-1 tenía cuatro integrantes, por lo cual en el libro de control se tenía la siguiente información:

Folio 2005	Ls
00001000	01
00001000	02
00001000	03
00001000	04

Pero en la ENNViH-2, el miembro con el ls "03" se desprendió del hogar original y formó su propio hogar de dos integrantes, para este nuevo hogar el folio quedó integrado de la siguiente forma.

Folio 2005	Ls
00001003	01
00001003	02

Asimismo, en la ENNViH-2, se construyó una variable llamada pid_link que es un indicador individual único que se mantiene constante a través de las diferentes rondas, el cual nos permitirá hacer el pegado de información individual a través de las diferentes rondas. La forma como se construyó el pid_link fue uniendo el folio más el ls que el individuo tenía en la ENNViH-1 sin importar si no se encontraban ya en este hogar. Para los individuos nuevos que se entrevistaron en esta ronda, el pid_link se construyó uniendo el folio y ls de la ENNViH-2.

Siguiendo con el mismo ejemplo anterior y suponiendo que al hogar original se incluyó un nuevo individuo y que el individuo con el ls = 3 se desprendió del hogar y formó su propio hogar el pid_link de los individuos quedó conformado de la siguiente forma en los respectivos hogares:

En el hogar original:

Folio 2002	Ls 2002	Folio 2005	Ls 2005	Pid_link	Ls01a
00001000	01	00001000	01	0000100001	1
00001000	02	00001000	02	0000100002	1
00001000	03	00001000	03		3
00001000	04	00001000	04	0000100004	1
-	-	00001000	05	0000100005	4

En el hogar nuevo:

Folio 2002	Ls 2002	Folio 2005	Ls 2005	Pid_link	Ls01a
00001000	03	00001003	01	0000100003	4
-	-	00001003	02	0000100302	4

Como se observa, el pid_link de los individuos nuevos que se agregaron al hogar original y al nuevo, quedó conformado por el folio más el ls de la ENNViH-2 ya que en la ronda anterior esos individuos no se encontraban en la muestra y por lo tanto no tenía ningún folio ni ls asignado. Para el hogar nuevo formado por el individuo con ls = 3, que fue el que se desprendió del hogar, en su hogar nuevo tiene su pid_link conformado con el folio y ls que tenía originalmente en la ronda anterior.

En la sección c_ls (Lista de los miembros del hogar) del libro C (Control), podemos identificar que individuo ya no es miembro del hogar mediante la variable ls01a (¿Sigue viviendo en este hogar?). Para los individuos que se desprendieron y fueron localizados en otro hogar el pid_link en el hogar original se les dejó en blanco, como se puede observar en el ejemplo anterior para el caso del hogar original, ya que en ese hogar no hay ninguna información de ellos.

Cuando el nivel de observación es diferente que el hogar o individual, usualmente es porque la base de datos fue recolectada como parte de una red, la cual es un conjunto

de preguntas que son repetidas por una serie de eventos o temas. Por ejemplo, en la sección salud y planificación familiar (SP) del libro C, cada observación corresponde a un determinado tipo de proveedor de salud o clínica y hay múltiples observaciones por hogar. En esta base de datos, la combinación del folio y la secuencia identifican a cada observación. La variable que define los temas o eventos en cada una de las bases de datos se llama secuencia.

En algunos casos, en los cuales los cuestionarios son individuales, existen bases en las cuales la información también se encuentra repetida por eventos o temas como es el caso del la sección de consulta externa (CE) del libro 3B, en esta base, para lograr tener un identificador único de cada una de las observaciones del segundo archivo (iiib_ce1), se necesita usar las variables folio, ls y secuencia.

5.3. Archivos de Cuestionarios de Hogares y de manuales de codificación.

- Existen doce archivos de cuestionarios, uno para cada libro de hogares (hh05q_byy); donde la letra q indica que esos archivos corresponden a los cuestionarios.
- Existen doce archivos de manuales de codificación. Uno para cada libro de hogares (hh05cb_byy), donde las letras cb indican que esos archivos se refieren a manuales de codificación.
- Un archivo zippeado contiene los doce archivos de los cuestionarios de hogares (hh05q_all).
- Un archivo zippeado contiene los doce archivos de los manuales de codificación (hh05cb_all).

Los archivos de los cuestionarios y de los manuales de codificación se encuentran en formato PDF.

5.4. Archivos Comunitarios

Al igual que en la encuesta de hogares, los archivos comunitarios incluyen: los archivos de la base de datos, los archivos de los cuestionarios y los archivos de los manuales de codificación. Los archivos comunitarios se presentan en la misma forma que los archivos de la encuesta de hogares:

- Existe un archivo zippeado que contiene todos los archivos de la encuesta comunitaria (loc05dta_all).
- Por su parte, dos archivos separados contienen la información de cada libro comunitario (loc05dta_byy), cada archivo corresponde a un cuestionario comunitario. Las letras yy indican el nombre de cada libro: El libro de características de la localidad (CC), el libro de precios de la localidad (p).

Los archivos de los cuestionarios y de los manuales de codificación se encuentran en formato PDF.

Todos los archivos tanto de la base de datos como demás documentación han sido comprimidos en archivos zippeados para facilitar su descarga. Con el fin de

descomprimir cada archivo zippeado es necesario tener instalado el software WinZip en su computadora. El software WinZip puede ser descargado gratis de <http://www.winzip.com/downwz.htm>.

6. Combinación de archivos

Para su mejor manejo, los datos de la ENNViH-2 son almacenados en archivos separados (por ejemplo, la sección ii_ah del libro 2 contiene información acerca de los activos y la sección ii_in, también del libro 2, contiene información acerca de los ingresos no laborales). Para el análisis de la base de datos de la ENNViH-2 usualmente se requiere hacer combinaciones entre los diferentes archivos de datos. La manera en que los datos pueden ser combinados depende de la naturaleza del análisis que se quiera llevar a cabo. En esta sección se discutirá algunas de las maneras para combinar archivos de datos.

6.1. Concatenando datos.

Algunas veces el análisis puede necesitar un conjunto de observaciones que se encuentran en dos diferentes archivos. Por ejemplo, en la sección iii_a_ata del libro 3A y la sección v_atn del libro 5, ambas secciones contienen datos sobre la asignación de tiempo. Los datos de la sección iii_a_ata se refiere a la asignación del tiempo de los adultos y la sección v_atn a la asignación del tiempo de los niños. Las variables para los adultos comienzan en ata, mientras que las variables para los niños en atn. En ambos casos, la información se refiere a lo mismo. Siendo así, podría ser interesante combinar estas dos variables para los dos grupos de edades. Para hacer este ejercicio es necesario emplear el comando APPEND en STATA o el comando SET en SAS. El archivo resultante contendrá tanto las observaciones de los niños como las observaciones de los adultos. Hay que notar que si el nombre de las variables es diferente entre los archivos que se quieren concatenar, las variables de alguno de los archivos podrían ser renombradas con el fin de poder hacer el match con el otro archivo.

En principio, diversos archivos pueden ser concatenados. Como regla general, cuando se usen datos de los libros 3A, 3B y 4, se debe verificar si alguna sección fue incluida en el libro Proxy, de esta manera los datos de encuestados que fueron respondidos por el mismo encuestado pueden ser combinados con datos recolectados en el libro Proxy para otros individuos.

La siguiente tabla muestra combinaciones adicionales. Se debe tomar en cuenta que puede ser posible que algunos archivos requieran ser reestructurados antes de que éstos sean combinados teniendo en cuenta los diferentes niveles de observaciones. También para algunos archivos se necesitará renombrar algunas variables.

Tabla 4. Archivos de hogares que permiten la concatenación		
Tema	Archivos	Tipo de informante
Activos	ii_ah y iii_a_ah	Informantes del libro 2 y libro 3A

Crédito	ii_crh y iiib_cr	Informantes del libro 2 y libro 3A
Ingreso no laboral	ii_in y iia_iin	Informantes del libro 2 y libro 3A
Violencia y crimen	ii_vlh y iia_vli	Informantes del libro 2 y libro 3A
Historia educativa	iii_ed y v_edn	Adultos y niños
Decisiones laborales	iii_tb y v_emn	Adultos y niños
Asignación del tiempo	iii_ata y v_atn	Adultos y niños
Consulta externa	iiib_ce y v_cen	Adultos y niños
Estado de salud	iiib_es y v_esn v_he y v_esn	Adultos y niños
Hospitalización	iiib_hs y v_hsn	Adultos y niños
Auto-tratamiento	iiib_ats y v_autn	Adultos y niños
Habilidades cognitivas	ea_eca y en_ecn	Adultos y niños

6.2. Combinación de archivos a nivel individual, hogar y comunitario.

Una de las formas más básicas para hacer combinaciones entre archivos involucra el cruce de archivos para una persona dada (a nivel individual). Si ambos archivos contienen datos en el mismo nivel de observación, el cruce será “uno a uno”

6.2.1. Combinación de dos archivos a nivel individual.

Supongamos que el objetivo del análisis es crear un archivo que contenga información acerca de la educación y las medidas de talla y peso de los individuos. El archivo iii_ed (adultos) del libro 3A y el archivo v_edn (niños) del libro v contienen información acerca de la educación de los individuos. El archivo s_sa del libro S contiene información acerca del peso y talla de los informantes. Los tres archivos contienen una observación por cada individuo. Para crear el nuevo archivo ordenamos cada uno de los tres archivos por folio y ls y posteriormente hacemos el merge por folio y ls.

Por ejemplo, supongamos que el investigador quiere crear un archivo con el peso del niño actualmente y el peso del niño al momento del nacimiento. El archivo s_sa del libro S contiene información acerca del actual peso, mientras que el archivo iv_he2 contiene el peso al momento del nacimiento. Tanto en s_sa como en iv_he2 hay una observación por individuo. Sin embargo, en s_sa las observaciones son identificadas usando folio y ls y en iv_he2 las observaciones son identificadas usando el folio y he06 (recordar que en la sección HE del libro 4 pregunta acerca de la historia de embarazo de las mujeres y se identifica cada embarazo con el identificador personal ls para cada hijo que es miembro del hogar). Primero es necesario quitar a los adultos de s_sa (recordar que este archivo contiene información acerca de los niños y de los adultos). Esto se puede hacer borrando a todas aquellas observaciones que tengan edades mayores al límite deseado. Entonces será necesario cambiar el nombre de he06 por ls

en iv_he2 (o ls por he06 in s_sa), ordenar los dos archivos por folio y ls y hacer el merge por folio y ls. Es importante tener en cuenta que si el universo de observaciones en s_sa es mayor al universo de observaciones en iv_he2, habrá observaciones que no coincidirán y por lo tanto estas observaciones deberán ser borradas.

6.2.2. Cruzando dos archivos a nivel hogar.

Para dos archivos a nivel hogar, tales como ii_in y ii_inr del libro 2, que contengan datos de las características del hogar, ordenamos cada archivo por folio y hacemos el merge por folio.

Cruzando dos archivos a nivel infraestructura de la comunidad de observación. Para dos archivos a nivel comunitario, tales como esc_dg y esc_de (características de la escuela) ordenamos cada archivo por folio y hacemos el merge por folio.

Cruzando dos archivos a nivel comunitarios. Supongamos que el investigador desea crear un archivo con información de las características generales de la escuela contenidas en esc_dg y las características de los pequeños proveedores de salud contenidas en peq_gp2 para cada comunidad. En este caso, antes de hacer el merge entre estos dos archivos, será necesario hacer el merge por folio de cada archivo con la portada de su correspondiente libro. Esto es importante debido a que queremos agregar una nueva variable id_loc, la cual es el identificador de cada comunidad y será la variable usada para el merge. Una vez teniendo esto para cada archivo se ordenará cada archivo por id_loc y el merge se hará por id_loc.

6.3. Combinación de archivos a nivel individual y hogar.

Frecuentemente, será necesario realizar cruces entre distintos archivos que no están en un mismo nivel de observación. Para este tipo de ejercicio algunas veces se necesitará reestructurar los diferentes archivos. Cuando se piensa acerca de la manera en que se pueden hacer los cruces entre los distintos archivos de la ENNViH-2 resulta de utilidad determinar si las variables que se identifican en uno de los archivos son un subconjunto de las variables identificadas en otro archivo.

Por ejemplo, supongamos que necesitamos crear un archivo que contenga información acerca del nivel de educación y de los activos del hogar de los adultos encuestados. La identificación de las variables del archivo de educación (iiia_ed) son folio y ls, pero la variable de identificación acerca de la información de los activos del hogar (ii_ah) es solamente el folio. En este caso, debido a que la variable de identificación en ii_ah es un subconjunto de las variables de identificación en iiia_ed, se podría simplemente hacer el merge mediante la variable folio. Esto es posible si las variables de identificación se encuentran a nivel de los individuos en iiia_ed. Esto dará como resultado una base de datos a nivel individual donde la información acerca de los activos del hogar es repetida para todos los individuos que son miembros de un mismo hogar.

Ahora supongamos que necesitamos hacer un merge (cruce) entre el archivo de educación (libro 3A) y el archivo acerca de crimen y victimización individual del libro

3A. Las variables de identificación en `iiia_ed` son folio y `ls`. Las variables de identificación en `iiia_vli1` son folio, `ls` y `SECUENCIA`. En `iiia_vli1` un individuo tendrá tantos registros como asaltos o incidentes violentos haya tenido. El archivo que queremos construir se creará de la siguiente manera:

Primero, dado que las variables de identificación en `iiia_vli1` son un subconjunto de las variables de identificación del archivo `iiia_ed`, se podría simplemente hacer un pegado por folio y `ls`. Esto dará como resultado una base que contenga tantos registros como incidentes violentos que haya reportado cada individuo. Cada registro contiene información acerca del nivel de alfabetización del individuo e información sobre un hecho violento particular.

La segunda opción es reestructurar el archivo `iiia_vli1` de tal manera que quede organizada a nivel individual más que a nivel de incidentes violentos reportados. En este caso, las variables de identificación serán folio y `ls`. Esto podría involucrar la creación de un archivo que contenga las variables, por ejemplo, `vli(01) – vli(15)` para victimización 1, `vli12_1_1`, `vli12_1_2`, ..., `vli12_1_15` (en este caso es 15 porque el máximo número de incidentes violentos que un individuo reporta es 15). Este archivo tendrá más variables que `iiia_vli1` pero tendrá un menor número de observaciones. Si los datos de `iiia_vli1` son reestructurados para hacer el archivo a nivel individual, el archivo resultado del pegado por folio y `ls` con el archivo de `iiia_ed` estará a nivel individual y contendrá información de alfabetización y toda la información acerca de la victimización.

Finalmente, habría que mencionar que la reestructuración de los archivos de datos para convertirlos a un nivel diferente del original se puede hacer relativamente con mayor facilidad en STATA con el comando `reshape` o en SAS con el comando `proc transpose`

6.4. Combinación de datos de hogares con datos comunitarios.

Cuando queremos combinar datos de la encuesta de hogares con datos de la encuesta comunitaria surgen dos casos básicamente: 1) datos de hogares con datos a nivel comunidad y 2) datos de hogares con datos a nivel servicio de infraestructura.

Combinación de datos de hogares con datos a nivel comunidad. Supongamos que se requiere crear una base de datos con información acerca del estado de salud de los niños –contenida en la sección `v_esn` del libro 5- y los precios de los medicamentos de los niños contenida en la sección `pre_prf2` del libro de precios en la comunidad. Lo primero que tendríamos que hacer para hacer el pegado entre estos archivos sería hacer un primer pegado del archivo `v_esn` con la portada del libro C por folio, esto con el propósito de agregar al archivo el identificador de la localidad `id_loc` (esta variable se encuentra en la portada del libro C). Del mismo modo, se tienen que hacer un merge del archivo `pre_prf2` con la portada del libro de precios en la comunidad (la variable `id_loc` se encuentra en la portada de los libros comunitarios y del libro C de hogares) por folio. Una vez teniendo la variable `id_loc` en los dos archivos, podemos hacer el merge entre los archivos por la variable `id_loc`.

7. Usando información de la ENNViH-2 con la ENNViH-1.

Esta sección provee una guía para usar las bases de datos de la ENNViH -1 y la ENNViH-2 y para obtener una base de datos longitudinal de los hogares e individuos.

7.1 Diferencias en el identificado de la ENNViH-1 Y ENNViH-2.

En la ENNViH-1 el identificador del hogar llamado folio, está conformado por 8 dígitos, de los cuales los últimos dos dígitos son ceros.

$$\frac{x\ x\ x\ x\ x\ x}{\text{FOLIO}} \quad \frac{00}{\text{terminación}}$$

En la ENNViH-2 el folio esta conformado igualmente de 8 dígitos pero a diferencia de la ENNViH-1 el folio puede tener una terminación diferente de cero. Si el hogar entrevistado es el hogar original, el folio es el mismo y los dos últimos dígitos serán ceros. Pero si un individuo original de la ENNViH-1 se desprende del hogar y forma su propio hogar, el folio de este individuo estará conformado por los primeros seis dígitos de su hogar original y los últimos dos dígitos se cambiarán por el identificador individual del individuo en la ENNViH-1.

$$\frac{x\ x\ x\ x\ x\ x}{\text{FOLIO}} \quad \frac{x\ x}{\text{ls}}$$

Por ejemplo, consideremos el hogar con el folio "00001000" en la ENNViH-1. En la ENNViH-2, un hijo de la familia se desprendió del hogar y formo su propia familia, el individuo que se separó del hogar en la ENNViH-1 tenía el identificador individual llamado ls "03". EL folio para las respectivas familias quedaría de la siguiente manera.

Folio	Tipo de hogar	Observaciones
"00001000"	Hogar original	Es el mismo hogar que se entrevisto en la ENNViH-1
"00001003"	Hogar Nuevo	Hogar que formo el individuo con el ls "03" del hogar anterior.

Debido a los desdoblamientos que hubo en los hogares originales en la ENNViH-2, cuando se formó un nuevo hogar el identificador individual del individuo también cambió y no es el mismo ls que el individuo tenía en el hogar original.

Siguiendo con el mismo ejemplo de la sección anterior, el individuo que formó un nuevo hogar, no tiene el mismo ls que tenía en el hogar original.

Tipo de hogar	Folio 2005	Ls 2005
Hogar original	"00001000"	"03"
Hogar Nuevo	"00001003"	"01"

Para tener un identificador individual único en ambas bases se creó un identificador llamado pid_link el cual permite identificar al mismo individuo en ambas rondas. El pid_link está conformado con el folio y ls original de los individuos, esto es, el folio y ls que les corresponde en la ENNViH-1, sin importar que en la ENNViH-2 se encuentren

en un hogar diferente a su hogar original. El pid_link está conformado por 10 dígitos, los primeros 8 dígitos son los mismos dígitos del folio y los últimos dos, son los correspondientes al ls.

$$\frac{\text{X X X X X X X X}}{\text{Folio}} \quad \frac{\text{X X}}{\text{Ls}}$$

Para los individuos nuevos que se encuentran en hogares originales o en hogares nuevos, el pid_link de estos individuos, estará conformado por el folio del hogar en el que se encuentren y su respectivo ls.

Por ejemplo, siguiendo con el mismo folio que hemos estado usando, y suponiendo que en el hogar original, cuando se llevó a cabo la ENNViH-2, se encontró en el hogar un nuevo miembro, el desdoblamiento de un miembro, el cual formó su nuevo hogar que está conformado por dos individuos. El pid_link para cada uno de estos individuos quedaría conformado de la siguiente forma.

Tipo de individuo	Folio ENNViH-2	Ls ENNViH-2	Folio ENNViH-1	Ls ENNViH-1	Pid_link
Jefe del hogar	"00001000"	"01"	"00001000"	"01"	"0000100001"
Nuevo miembro	"00001000"	"04"	-	-	"0000100004"
Miembro desprendió del hogar	"00001003"	"03"	"00001000"	"03"	"0000100003"
Nuevo miembro en hogar nuevo	"00001003"	"02"	-	-	"0000100302"

7.2 Pegando información de la ENNViH-1 y la ENNViH-2.

Antes de hacer el pegado de información entre la ENNViH-1 y ENNViH-2, se necesitan tener las bases generadas con toda la información que se requiere de cada ronda de acuerdo a la explicación dada en los puntos anteriores. Una vez que tenemos las bases listas, la información puede ser pegada a nivel individual, hogar y comunitario.

7.2.1 Combinación de datos a nivel hogar.

Para hacer el pegado de información entre ambas bases a nivel hogar, existen varias opciones con las cuales se puede hacer el pegado. Si se quiere solo pegar la información de ambas bases a los hogares originales de la ENNViH-1, el pegado se tiene que hacer con el folio. De esta forma solo se le estará pegando la información a los hogares que se encuentran en la ENNViH-2 que son los mismos hogares originales que se encuentran en la ENNViH-1.

Si lo que se quiere hacer es pegarle a los hogares de la ENNViH-2 la información del hogar al que pertenecían en la ENNViH-1 sin importar si son hogares originales o nuevos, entonces se tendrá que generar una variable que sean los primeros 8 dígitos del pid_link y llevar a cabo el cruce. Se usa el pid_link y no el folio, ya que existen

individuos dentro de la base que son individuos originales pero que formaron un hogar nuevo con otro individuo que también es individuo original de nuestra muestra y al formar el nuevo hogar, se va a usar el folio de uno de ellos dos para formar su nuevo folio. La forma de poder rastrear el hogar del cual fueron desprendidos es a través del `pid_link`.

La forma como se construye en stata el folio original a partir del `pid_link` es de la siguiente forma:

```
gen str8 var = substr (folio, 1,8)
```

A diferencia de la ENNViH-2, el folio en la primera ronda se encuentra numérico y para poder llevar a cabo el pegado de las dos bases se necesita tener en el mismo formato ambos folios. Para volver el folio de la ENNViH-1 en string usando stata se realiza con la siguiente instrucción:

```
gen str8 var1=string(folio, "%08.0f")
```

Donde `var1` es la nueva variable que se está generando.

7.2.2 Combinación de datos a nivel individual.

Si el pegado se quiere hacer a nivel individual entonces el pegado se tiene que hacer con el `pid_link`. En las bases de la ENNViH-1 el `pid_link` no está generado, por lo cual solo se tiene que construir como se explica a continuación para posteriormente hacer el *merge* con esta variable.

Comandos en Stata para generar el `pid_link` en la ENNViH-1:

```
gen str8 var1 = string (folio, "%08.0f")
gen str2 var2 = string (ls, "%02.0f")
gen pid_link = var1 + var2
```

La generación de las variables `var1` y `var2` son para generar el folio y `ls` en string de 8 y 2 dígitos respectivamente ya que en la ENNViH-1 se encuentran en formato numérico y cuando el folio tiene menos de los 8 dígitos se debe de llenar con tantos ceros sean necesarios a la izquierda para conformarlo de 8 e igualmente con el `ls`, cuando el `ls` sea de un dígito, se le antepone un cero para que quede conformado de los dos dígitos que debe de ser.

Folio ENNViH-1	ls ENNViH-1	Var1	Var2	Pid_link
1000	1	"00001000"	"01"	"0000100001"

7.2.3 Combinación de datos a nivel comunidad.

Si las bases de datos a combinar se encuentran a nivel comunidad, el pegado de ambas bases tiene que realizarse usando la variable folio que se encuentra en las bases de datos comunitarias.

8. Números de preguntas y nombre de las variables.

La mayoría del nombre de las variables de la ENNViH-2 corresponde a los números de las preguntas de la encuesta. Por ejemplo, los nombres de las variables de la sección ED (Educación) comienzan con ED y con el número de la pregunta correspondiente. En los cuestionarios de la ENNViH-2 se trató que el número de las preguntas correspondiera al número y nombre de las preguntas de la ENNViH-1. Si alguna pregunta se adicionó o se cambió en la ENNViH-2, generalmente se le agregó a esta pregunta la letra "a" o "b".

8.1. Valores Missing y códigos especiales.

Para las variables numéricas, un punto (.) significa un dato faltante (missing). Para las variables de texto (string) un espacio significa un dato faltante (missing). Los valores missing contenidos en algunas variables puede ser resultado de no respuesta debido a los propios saltos o filtros contenidos en el mismo cuestionario. Por ejemplo, la pregunta ah01 del libro 3A le indica al entrevistador que verifique si el encuestado ya respondió el libro 2, si es así, a esta persona no se le pregunta esta sección y se pasa a la siguiente sección. Si la respuesta es si (ah01=1) el encuestado preguntará la sección AH completa. Otros casos en donde pueden aparecer valores missing son en los que los encuestados no quieren o no saben responder la pregunta. En este caso existen códigos especiales cuya terminación puede ser; 5, 7, 8 o 9 según el caso.

Estos códigos especiales pueden tomar estos valores dependiendo de cada pregunta y del espacio asignado para la respuesta:

5, 95, 995 ó 9995 = fuera de rango, la respuesta no se puede anotar debido a que el espacio del cuestionario es insuficiente.

7, 97, 997 ó 9997 = El encuestado no quiso responder la pregunta

8, 98, 998 ó 9998 = El encuestado no conocía la respuesta

9, 99, 999 ó 9999 = La pregunta no aplica.

8.2. Tipos de respuestas.

La mayoría de las preguntas de la ENNViH-2 requiere como respuesta un valor numérico. Sin embargo, hay casos en donde pueden existir respuestas abiertas. Para las preguntas cuya respuesta debe ser numérica existe un máximo número de dígitos permitidos, cuando alguna respuesta supera este máximo de dígitos, el entrevistador deberá usar el correspondiente código especial.

Algunas preguntas permiten múltiples respuestas. Por ejemplo, ¿Quién en tu familia decide acerca de la educación de los hijos? En este caso, se han especificado respuestas alfabéticas. Una diferente variable fue creada para cada opción así que las nuevas variables contienen 1 si el encuestado responde afirmativamente y sin valor cuando responde negativamente.

Para variables categóricas, el cuestionario provee las diferentes opciones para cada respuesta. Los manuales de codificación también proveen la distribución de las respuestas. Para variables categóricas también se proporciona la distribución de frecuencias. Para este tipo de variables donde respuestas múltiples son permitidas, los

manuales de codificación indican el número de encuestados que dio cada respuesta. Si fuera posible tener muchas combinaciones de respuestas, el manual de codificación no da la distribución de todas las respuestas.

9. Características especiales de la información de los hogares de la ENNViH-2.

En esta sección se presentan características particulares de los datos de hogares de la ENNViH-2 que pueden afectar el análisis.

9.1. Información simétrica

En dos secciones del libro IIIA; activos (AH) y decisiones del hogar (DH), esposos y esposas dan la misma información. Es decir, los esposos responden las preguntas relacionadas con el y con el resto de la familia (incluyendo su esposa). Del mismo modo, la esposa responde las mismas preguntas; es decir, responde para el caso de ella y para el caso de los demás integrantes del hogar (incluyendo su esposo). Estos datos permiten llevar a cabo comparaciones acerca de lo que piensan las parejas de sí mismos y de los otros miembros del hogar.

La sección AH (activos) del libro IIIA únicamente es respondida por miembros del hogar con pareja, cuya pareja ya había dado información acerca de los activos del hogar preguntada en el libro II. En un hogar, si dos individuos están casados, la información de sus activos podría ser la misma y la información de una persona puede ser usada para cubrir posibles respuestas omitidas de la otra persona. Similarmente, en la sección DH del libro IIIA los individuos casados describen la manera en que las decisiones dentro del hogar son tomadas. La información de ambos informantes puede ser usada para verificar consistencias y en caso de falta de información de alguno de ellos se podrá complementar la información.

Lo mismo sucede en la sección TH (hermanos) del libro IIIB, donde se recolecta información detallada acerca de hermanos no residentes en el hogar. Nosotros recogemos información acerca de las características de los hermanos no residentes del hogar de todos los miembros del hogar.

9.2 Información duplicada.

Existe información que fue preguntada en varias secciones de los cuestionarios. En la mayoría de los casos, el informante fue una fuente de información y un informante Proxy (o información pre impresa) fue otra. Por ejemplo, en el libro 2, la sección de activos del hogar (AH) contiene información que también fue preguntada en los libros individuales. Aunque pueda resultar más fácil usar la información agregada del libro del hogar, la información que proviene de los libros individuales muy probablemente es más acertada o puede ser utilizada para complementar la información de la primera fuente.

Edad. La información referente a la edad se pregunta en la sección LS del libro C (generalmente un miembro del hogar indica la edad de todos los miembros del hogar) y en las portadas de cada uno de los libros individuales. Nosotros no corregimos posibles inconsistencias de edad entre lo reportado en el libro C y en las portadas de los libros individuales.

Sexo. La información sobre el género del individuo se pregunta en la sección LS del libro C, en la sección SA del libro S y en las portadas de los libros 3A y 3B. De manera similar que en el caso de edad, nosotros no corregimos posibles inconsistencias entre lo reportado en el libro C y lo registrado en las portadas de los libros individuales.

Estado Civil. El estado conyugal fue preguntado en la sección LS del libro C y en las portadas de los libros 3A, 3B y IV. En la limpieza de los datos, intentamos cerciorarnos de que el estado civil registrado en la lista LS era comparable con el estado civil registrado en las portadas de los libros.

Nivel de educación. La sección LS del libro C se reporta el máximo nivel y grado educativo alcanzado. Para algunos individuos esta información se vuelve a preguntar en el libro 3A y 5.

Ingreso laboral y no laboral. La sección TB del libro 3A pregunta acerca de los ingresos laborales. Estos temas también se preguntan en el libro Proxy. En la sección IN del libro 2 se pregunta acerca del ingreso no laboral a nivel hogar y en la sección IIN del libro 3A se pregunta a nivel individual. La información a nivel individual es preferida, pero los datos agregados a nivel hogar son útiles para analizar el ingreso total del hogar en caso de que falte esta información de algún miembro del hogar.

Estatus de supervivencia de los padres. La lista de miembros del hogar se encuentra registrada en la sección LS, que contiene el número de línea del padre y madre de cada uno de los miembros del hogar (ls06 y ls07). Se utilizaron códigos especiales para registrar si los padres viven pero no son miembros del hogar o si el padre o la madre falleció. En el libro 3B, en la sección TP (padres) se pregunta explícitamente acerca de si el padre o madre está vivo o muerto.

9.3. Relación entre los miembros del hogar.

La ENNViH-2 contiene información acerca de las relaciones familiares, particularmente entre esposos y esposas y entre padres e hijos, de tal forma que es posible identificar quien es cónyuge de quién y quién es hijo de quién. La información no se limita a miembros del hogar, sino que también cubre a relaciones de no residentes.

Relación de Parentesco en el roster LS (Libro C)

La sección LS da información completa acerca de la relación de parentesco entre los miembros del hogar actual. La información dada por la lista de miembros del hogar LS es como sigue:

Tabla 5. Relación de parentesco		
Variable	Información	Observaciones
Ls05	Quién es el jefe del hogar en la ENNViH-2 y qué parentesco tienen los demás miembros del hogar con el jefe del hogar.	Esta información indica el parentesco que tiene cada miembro del hogar con el jefe del hogar. Por ejemplo, si las personas con ls 3 y 4 fueran hijos del jefe del hogar ellos serían hermanos o medios hermanos. Si la persona 5

		fuera la madre del jefe del hogar y la persona 3 fuera el hijo del jefe del hogar, la persona 5 sería con toda seguridad la abuela de la persona 3. En otros casos la información no es definitiva. Por ejemplo, si las personas 6 y 7 fueran nietos del jefe del hogar, ellos podrían ser hermanos o primos.
Ls06	Número de línea (LS) del padre del miembro del hogar.	Esta variable permite vincular a un individuo con sus padres mediante su id individual. Por ejemplo, si la persona X tiene un ls06=1 y ls07=2, su padre es el jefe del hogar (debido a que el jefe del hogar siempre tiene un LS=1) y su mamá es la persona con el LS=2. Esta variable indica quién es la persona que dentro del hogar cuida a los niños menores de 15 años.
Ls07	Número de línea (LS) de la madre del miembro del hogar.	
Ls08	Número de línea (LS) de la persona que cuida a los niños menores de 15 años.	

Identificación de hijos en otras secciones

Las secciones HE1 y HE2 del libro IV contienen información acerca de la historia de embarazo de todas las mujeres de entre 14 y 49 años de edad. La información recolectada incluye las características de los hijos en el momento del nacimiento y los cuidados de salud que antes y después del nacimiento tuvo la madre.

Estas secciones identifican cada nacimiento con el número de identificación personal (LS) para cada hijo que es miembro del hogar. Esto permite cruzar información de los hijos nacidos con otra información de los mismos hijos contenida en la encuesta. Los hijos cuyo ls es desconocido fueron codificados con un "99".

Hijos no miembros del hogar. La sección THI del libro 3B da información detallada acerca de los hijos no miembros del hogar. No se recopiló información sobre las características de los hijos (miembros del hogar) si la información ya había sido dada por algún otro informante. Por ejemplo, si la persona X (un hombre) está casado con Y (mujer) y no tuvo otras esposas, sus hijos son también hijos de la persona Y. En este caso, la información provista por la persona Y acerca de sus hijos es utilizada como la información de los hijos de la persona X.

Si la persona X tuvo otros hijos con otra mujer, la persona X responderá las preguntas acerca de sus otros hijos.

Hermanos no miembros del hogar. La sección TH (hermanos) en el libro 3B da información detallada acerca de los hermanos no miembros del hogar. En este caso, y contrario a lo que sucede en el caso de hijos no miembros del hogar, se recolectó información acerca de las características de los hermanos no miembros del hogar de todos los miembros del hogar, a pesar de que la información de algún hermano no miembro del hogar haya sido dada por otro miembro del hogar.

Para calcular el número total de hijos, hermanos o padres de los informantes u obtener información acerca de las características de estos parientes, es necesario cruzar la información de diferentes secciones. La siguiente tabla muestra como hacer el cruce de información.

Tabla. 6 Fuentes de información para los familiares cercanos de los informantes

Hermanos	
En el hogar	<ul style="list-style-type: none"> • Para un jefe del hogar, usar la variable Is05 de la sección LS del libro C para identificar a miembros del hogar quienes sean hermanos y hermanas. Para cualquier otro miembro cuyo padre o madre sea miembro del hogar, verificar la lista de miembros del hogar (LS) para otros individuos del hogar quienes identifican a sus padres en Is06 y Is07 de la sección LS del libro C. Para un miembro del hogar no jefe del hogar cuyos padres no son miembros del hogar la información de los hermanos es nuclear. • Una vez que se identificaron a los hermanos y hermanas de los miembros del hogar, es posible consultar su información en los libros individuales.
Fuera del hogar	Usar la información de las características de los hermanos no miembros del hogar de la sección THI del libro IIIB.
Hijos	
De informantes mujeres	Para cada hijo nacido, usar la información de la sección HE del libro 4. Para hijos vivos también se puede obtener información en el libro 5.
De informantes hombres	Para cada hijo de la pareja actual, usar la información dada por la esposa/cónyuge de la sección HE del libro IV y libro 5. Para hijos no miembros del hogar mayores de 15 años de una pareja previa usar los datos de la sección THI (Hijos) del libro 3B.
Padres	
En el hogar	Dependiendo de la información requerida, usar datos de los libros 3A y 3B de cada padre o de la sección LS del libro C.

Fuera del hogar	Usar la información de la sección TP (Padres) del libro 3B.
-----------------	---

10. Limpieza de la información de la ENNViH-2

Esta sección describe el proceso de sistematización de la información de la ENNViH-2. Uno de los procesos fundamentales para cualquier encuesta es la captura, sistematización y limpieza de la información recolectada por el equipo de entrevistadores. En el caso de la ENNViH, el proceso de la captura de la información se llevó a cabo en campo casi inmediatamente después de haber sido recolectada.

Una vez concluida la fase de diseño de cuestionarios, se diseñó el sistema de captura de la información, CAFÉ por sus siglas en inglés (Computer Assisted Field Editing). El CAFÉ fue instalado en lap tops y el equipo de editores –quienes también se trasladaban a la localidad con el equipo de entrevistadores- eran los encargados de llevar a cabo la captura de la información.

De esta manera, una vez que los entrevistadores llenaron la información del hogar en los cuestionarios impresos en papel y que fueron revisados por el supervisor de campo, se entregaban al equipo de captura. El proceso de captura en campo permite por un lado, de que en caso de existir omisiones, errores o inconsistencias de la información en el momento de la captura, regresar a los hogares a completar y/o verificar estos errores o inconsistencias y por otra parte, hace que el proceso de limpieza de la información sea más eficiente y más rápido.

El equipo de edición incorporó y corrigió la información directamente en las lap tops, usando el sistema de captura de información diseñado para detectar errores u omisiones. Por ejemplo, si el género de un informante registrado en la lista LS no coincidía con el género registrado en la portada del libro 3A se genera un mensaje de error. El editor sería el responsable de discutir los posibles errores con el entrevistador. Con el sistema de captura en campo, algunos de los errores pueden resolverse fácilmente. En el caso del ejemplo anterior, el entrevistador puede recordar fácilmente el género de la persona encuestada, pues la entrevista se llevó a cabo en días pasados. Otros errores requirieron que el entrevistador regresara al hogar a verificar los posibles errores u omisiones directamente con el informante.

El CIDE y la UIA hicieron la limpieza de datos adicionales para corregir errores restantes y para hacer la información del dominio del público.

10.1 Sistema de búsqueda y proceso de limpieza. “Look Ups”

Para detectar y resolver errores más complejos, se puso en ejecución el proceso de limpieza llamado “look Ups” (LU). El LU implicó el uso de programas computacionales sofisticados que garantizaran la calidad de la información. Este proceso hace uso de información adicional contenida en los cuestionarios que no fue capturada de forma electrónica. Por ejemplo, una inconsistencia pudo ser generada porque se hizo una corrección inadecuada; la referencia a la anotación original del entrevistador resolvió la edición, de esta manera los datos pudieron ser corregidos.

Como ejemplos, el programa del LU puede verificar que los padres fueran tuvieran por lo menos 12 años más que sus hijos. El programa genera automáticamente una detección de este tipo de errores los cuales pueden ser corregidos de forma transparente y verificados utilizando los cuestionarios originales. Sin embargo, si el especialista no estaba muy seguro de cómo corregir los datos, los datos no fueron corregidos. Algunos problemas eran relativamente fáciles de corregir. Otros, como los filtros o saltos omitidos, no podían ser corregidos porque los datos no habían sido recogidos.

10.2 Limpieza especial para respuestas abiertas

Respuestas abiertas. Algunas preguntas de los cuestionarios permitían las respuestas abiertas. La limpieza de estas respuestas abiertas fue hecha por un equipo especialmente entrenado.

VARIABLES CON RESPUESTAS DE "OTRA, ESPECIFICAR". Este tipo de respuestas tienen lugar cuando una respuesta no correspondía a las opciones de respuesta previamente codificadas. En la limpieza de las respuestas "otras, especificar", fue necesario verificar las respuestas del texto y decidir si alguna respuesta se podría adicionar en una categoría existente, o si era necesaria la creación de una nueva categoría. La limpieza fue realizada por equipos especialmente entrenados en el CIDE y la UIA. Las nuevas categorías fueron creadas cuando alguna respuesta fuera substancialmente diferente a las respuestas precodificadas.

Por lo tanto, los usuarios tienen la opción de agregar los datos. Una desagregación más fina de los datos sería imposible si los nuevos códigos no hubieran sido creados. Estas remodificaciones serán tomadas en cuenta en la elaboración de instrumentos (cuestionarios) de futuras rondas de la ENNViH.

Fueron limpiadas tres tipos de respuestas de "otra, especificar":

- Preguntas simples que permiten una sola respuesta: Para esta clase de preguntas, las respuestas de "otras" fueron recodificadas en una nueva o existente categoría de respuesta. Por ejemplo, la pregunta 11 de la sección CV del libro C pregunta si el hogar desinfecta el agua que usan para beber. El cuestionario original permitió 5 categorías más la opción de "otra". Después del proceso de limpieza, una nueva categoría fue agregada: "compra el agua purificada".
- Preguntas donde se permiten respuestas múltiples: La opción de "otra" fue recodificada a una nueva o existente categoría, y se eliminó la opción de "otra". Por ejemplo, la pregunta GH02 de la sección GH del libro 3B pregunta qué tipo de bebidas prefiere tomar mientras come. El cuestionario original permite 4 categorías (agua, cerveza, soda y tequila/pulque) más la opción de "otra". Después del proceso de limpieza, dos nuevas categorías fueron agregadas (ron y las bebidas calientes) y la "otra" opción fue dividida en "otras bebidas alcohólicas" y "otras bebidas sin alcohol".
- Preguntas relacionadas con los artículos en una matriz. La limpieza de "otras" respuestas aquí pudo generar otro artículo (renglón) en la matriz. Si una nueva

categoría fue creada. Por ejemplo, la sección CA del libro 3b preguntó por los tipos de seguros que contaban los individuos. Esta sección se agregó el renglón de Seguro Popular que no existía en la ENNViH-1.

Al final del proceso de la limpieza de los datos, las observaciones que seguían teniendo la opción de "otro" están en categoría de texto. Estas variables no están disponibles al público.

Variables Numéricas. Algunas respuestas numéricas no cupieron en el espacio proporcionado. En estos casos, se capacitó a los entrevistadores para completar el espacio proporcionado con un código de 9 que terminaban en 5 ("falta de espacio") y registrar la respuesta correcta en la sección del cuestionario "observaciones" o en "otro, especificar". Si la justificación proporcionada por el entrevistador era válida, se añadieron campos numéricos para permitir la respuesta correcta y se substituyo "falta de espacio" por la respuesta correcta. No fue posible corregir todas las respuestas con este código, así que los códigos especiales a veces aparecen en los datos.

10.3. Sección de verificación.

En cada sección se hizo un esfuerzo para:

- Verificar el proceso del LU y determinar si algunos errores o inconsistencias restantes podrían ser corregidos
- Verificar las respuestas numéricas para la existencia de códigos especiales y verificar las variables con respuestas vacías o códigos de "no sabe".
- Crear o corregir las variables "_ 1", "_ 2" de modo que los códigos especiales fueran preservados y el número asociado a alguna variable contuviera sólo respuestas válidas. Las variables "_ 1" se asocian generalmente a un valor numérico e indican si la persona podía o no contestar a la pregunta
- Verificar que los "saltos" o "filtros" fueran seguidos correctamente y hacer las correcciones pertinentes en caso de que la omisión de datos no fuera por los saltos o filtros.
- Asignar los nombres y las etiquetas variables tan claramente como fuera posible.
- Verificar la existencia de observaciones duplicadas
- Encontrar y eliminar cualquier variable que pudiera permitir la identificación del informante.

10.4. Verificación de los identificadores a través de los libros.

Es esencial que los números de identificación (identificación) FOLIO y el LS estén asignados correctamente. Por lo tanto, se verificó rigurosamente la asignación de la identificación usando un especial programa que utilizó los mismos identificadores a

través de todas las secciones. Este programa verificó que la edad y el nombre de cada individuo estuvieran correctos en todas las diferentes secciones

10.5. Verificación de las unidades de medida.

En algunas preguntas la respuesta era numérica permitiendo alguna opción de medida. Por ejemplo, la pregunta CS10 del libro I (consumo del hogar) pregunta sobre la cantidad de productos comprados en los últimos 7 días, pudiendo tener como respuesta kilogramos, piezas, o alguna otra unidad. Los informantes pudieron haber proporcionado respuestas claramente equivocadas. Por ejemplo, si un individuo dijo que había comprado 10 kilogramos de pan, la unidad apropiada para esta respuesta es más bien 10 piezas en vez de kilos. De igual manera, si una mujer reportó haber tenido un aborto después de un embarazo de 11 meses, juzgamos que la unidad apropiada debió haber sido semanas en lugar de meses. Tales correcciones implicaron muy pocos casos.

10.6. Creación de variables y archivos.

Algunas variables y archivos de datos fueron creados con el fin de hacer su uso más fácil. Por ejemplo:

Las respuestas asociadas a la ocupación y sector de ocupación del libro 3A estaban codificadas, pero también se permitían respuestas abiertas. Las respuestas abiertas fueron codificadas posteriormente para hacerlos comparables con los códigos del CMO para la ocupación para las clasificaciones de la actividad. Estas nuevas variables se nombran `... cmo ' y son seguidas por el número de la pregunta correspondiente.

La información referentes a la edad y fecha de nacimiento, así como la información de educación, puede variar entre los libros, nosotros construimos nuestra "mejor conjetura" de estas variables para cada persona usando todos los datos de la ENNViH-1 lo cual ese encuentra en base de datos de BESTGUESS? Para el libro 4 (salud reproductiva) substituímos los nombres de los niños que fueron proporcionados por sus madres para el número correspondiente de la identificación de cada niño. Esto se podría hacer solamente para los niños que vivieron con sus madres y cuyas madres se encontraban entre 14 a 49 años de edad. Éste es el caso de he06 variable contenido en la sección HE1, y de he19a variable contenido en la sección HE2. Estas variables asignan un código del ` 99' para esos niños que no encontramos en la casa.